91602
new 82-10

COMPLIMENTARY COPY
Please Credit Ontario Hydro Research if
Reproduced in Whole or in Part

ontario hydro
research division

To: Dr. D. Mills
Director of Research

USE OF NEURAL NETWORKS FOR FISH IDENTIFICATION FROM SONAR ECHOES - PREPROCESSING AND NETWORKS IN PARALLEL

N. Ramani
Operations Research
Department

W.G. Hanson
Divisional Projects
Department

P.H. Patrick
Environmental Sciences
Department

Studies are currently under way in Ontario Hydro to investigate the use of sonar based methods for collecting environmental fisheries data. The work reported here is a part of these studies and involves the use of neural networks to identify the fish from their sonar echoes. The current phase of the work investigates the impact of preprocessing techniques and the use of networks in parallel on the generalization properties. It is found that significant improvements are possible using simple parallel combinations of three-layer perceptrons which have been trained using outputs from different preprocessors. In the test case studied, over 95 percent of the targets were identified correctly by the parallel combination.

1. INTRODUCTION

To address regulatory requirements, Ontario Hydro conducts periodic aquatic studies in the neighbourhood of all its generating stations in an attempt to determine the impact of its activities on fish populations. The Research Division is investigating the potential of sonar to replace, to some extent, the current netting practices, because it is safer, non-consumptive and less labor intensive. One key element of the proposed approach is the identification of the fish from their sonar echoes.

An earlier report (Ramani, et al, 1991b) outlined some of the preliminary work done on laboratory data at Ontario Hydro to explore the potential of neural networks for the identification task and concluded that the

job	file	date	report no.
74-664-090	810.0	November 14, 1991	91-254-K

LEGAL NOTICE

This report was prepared by Ontario Hydro and, while effort has been made to assure the accuracy and applicability of the material contained herein, neither Ontario Hydro nor any person acting on its behalf:

(a) makes any warranty, express or implied with respect to the use of any information, apparatus, method or process disclosed in this report or that such use may not infringe privately owned rights; or

(b) assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method or process disclosed in this report.

method is viable. This report summarizes the results of a study which expands on that work and examines the impact of the choice of the training/test sets, preprocessing methods, and the use of networks in parallel. Some of this work has been reported earlier by Ramani, et al (1991a) and Patrick, et al(1991).

2. DATA

The data used in this study were collected at Ontario Hydro's laboratories using a large semi-anechoic tank (approximately 15 m x 15 m x 1.5 m). A dual beam 420 kHz transducer (6 degrees and 15 degrees) with time varied gain was placed at one end of the tank, and the targets were approximately 6 m from the transducer. The test targets consisted of three types of debris (a stream of bubbles, a leaf, and a ping-pong ball) and four species of fish (brown bullhead, rainbow trout, sturgeon, and walleye). The air bubbles were created by using a wand, and the other objects were suspended using a monofilament transparent to the transducer. The four fish samples were suspended laterally with respect to the transducer. Two additional datasets were created by suspending the sturgeon frontally and at an angle of 45 degrees with respect to the transducer, making a total of nine targets.

Pulses 0.5 ms wide were transmitted every 0.5 s and the reflected signals were received through a wideband sonar detection module (developed at Ontario Hydro) and recorded in an FM recorder. The carrier signal is filtered out and the recorded signal consists only of the envelope representative of the target. Only the 6 degree data were used and forty eight signals from each of the nine targets were digitized at 60 kHz. Signal echoes 192 points long were retrieved from the digitized data using the transmitted pulse as a guide. Figures 1 and 2 show some sample echoes. It can be seen from Figure 1 that echoes from the different species can in fact look similar, whereas Figure 2 illustrates the variations in the echoes from a single species.

3. ANALYSIS

3.1 The Neural Network Models

The neural network approach is a biologically motivated statistical method for mapping between vector spaces. In recent years, the approach has been used successfully in many pattern recognition applications where a neural network maps an input pattern vector into an output decision vector. Appendix A presents an outline of the method.

All the networks in this study were multi-layer perceptrons trained using the back-propagation algorithm. The size of the first layer varied depending on the preprocessor. The number of units in the middle layer were varied from 5 to 30 units and the output layer consisted of 9 units corresponding to the nine targets. The weights of the networks were initialized to random values between -0.3 and 0.3. The learning rate and momentum factors (see Appendix A) were set at 0.01 and 0.6 respectively and training continued until the rms error on the training samples reached 0.05 (The signals were normalized to a range between 0 and 1

prior to the training so that the maximum possible error is 1.0). In most of the cases, this meant that all the training samples were learnt.

The dataset was divided into three parts (labelled A, B and C) and, except for the first set of networks (see Section 3.2), each net was trained on two parts and tested on the third. That is, each training set consisted of 32 signals per target and the test set consisted of 16 signals per target. The whole dataset is spanned by studying the three cases obtained by permuting the training and test sets.

3.2 Original Data - Choice of Training set

The first test examines the impact of the choice of the training set on the generalization properties of the trained network. One set of networks was trained on one subset (ie, the training set consists of 16 samples per target) and tested on the other two. A second set of networks was trained on two subsets (32 samples per target) and tested on the third subset. All the possible permutations of the training/test sets were studied, and the impact of the choice of the training set could thus be assessed by inspecting the results.

Tables 1 and 2 present the results for the two sets of networks. The overall performance of the networks in Table 2 (two subsets used for training) are definitely better than those in Table 1 (one subset used for training). Clearly, the variations in the individual samples are significant enough to require the additional samples for training. The rest of this study uses two subsets for training and one for testing.

The average success rate for the networks in Table 2 is around 80 to 85 percent and the best performance is obtained with a hidden layer size of 20 or 25. The performance is fairly consistent over all the three subcases (permutations of the training/test sets) although the generalization in the second subcase (trained using subsets A and B, and tested on subset C) seems to be the poorest.

3.3 Four Layered Nets

It is well known that one needs at least one hidden layer (ie, a total of at least 3 layers) to realize most complex mappings. However, the optimum number of layers depends on the problem under consideration. Several studies have tested the impact of four-layered networks and some have reported improvements. In the present case, however, no improvement results, as shown in Table 3. The overall success rate is about the same as in Table 2 for the three-layered networks.

3.4 A Wavelet Based Approximation

The next step in the analysis was the study of the impact of preprocessing techniques on the generalization properties of the networks.

The first preprocessing technique involves the approximation of the signal using an orthogonal multiresolution representation called a wavelet representation (Mallat, 1989). The signal is approximated at a spectrum of resolutions in an orthonormal basis. Studies are under way in an attempt to use the different resolutions to capture different features of the signals. The work reported here refers to only one resolution.

The filter used corresponds to an approximation based on cubic splines and uses an approximation at 1/4 resolution (with each signal having one fourth the number of points as those in the original, ie, 48 points).

Table 4 shows the results using this preprocessed set of signals as inputs to a neural network. There are marginal improvements in performance, but the training times were considerably lower (approximately one-half) compared to the net trained on the original data. It is surprising to see an improvement in performance when an approximation is used, and it appears that any loss of information is more than compensated by the reduction in noise. The overall success rate is again 80 to 85 percent and the second subcase has the poorest generalization.

3.5 A Peak/Trough Detector

The second preprocessing method is a simple peak/trough detector. Each signal is represented by 80 values representing ten peak and ten trough values and their locations for both the signal and its derivative. Table 5 shows the results. The nets based on this preprocessor performed better and the average overall success rate was about 85 to 90 percent. However, more interestingly, the networks are able to make the correct identification in many instances where the networks described in Section 3.4 failed. The best generalization is obtained in the second subcase whereas in Section 3.4 (using the wavelet based approximation), the second subset had the poorest generalization.

3.6 Networks in Parallel

There could be several reasons as to why the two sets of networks in the previous two sections generalize differently. One likely reason is that the preprocessors accentuate different features of the signal resulting in successes in different subsets of the input space. This suggests a third set of networks which somehow combine the successes of the previous two sets of networks. Several possible combinations were attempted and the most successful one is reported here. Note that typically, the guess of a network corresponds to the output node with the maximum activation. In the parallel network used here, outputs of two nets based on the two preprocessing techniques are averaged before computing the guess of the network.

Table 6 shows the results when each network described in Section 3.4 (Table 4) is combined in parallel with the corresponding network described in Section 3.5 (Table 5). The parallel networks clearly perform better than either of the above two sets of networks and success rates range around a very impressive 95 percent for all the three subcases. For this set of data, the combined decision surface is certainly a better discriminant than either of the two individual discriminant surfaces used alone. The idea of using networks in parallel seems to be very promising and work on this area is continuing.

4. CONCLUSIONS

This report has summarized a set of preliminary tests on the use of neural network based methods for the identification of fish from sonar echoes. It was found that sample variations were significant enough that nets trained on only one-third of the samples did not generalize adequately. Preprocessing did have a

noticeable impact on the results and tests on the use of other schemes should be worthwhile. The optimum number of neurons in the middle layer for a three-layered net was around 20 to 25 for this set of data and four-layered nets did not offer any improvements over three-layered nets.

The most striking result was that significant improvements in performance are possible with simple parallel combinations of two networks, which have been trained using the outputs of different preprocessors.

Overall, the tests based on the laboratory data continue to be encouraging. However, the project is still in early stages and further work involving other species of fish, fish at other different angles with respect to the transducer, the use of other preprocessing schemes and tests on field data (Mattagami River) are under way. Although this study has focussed on the high frequency (420 kHz) sonar echoes, the method will also be applicable to the detection of fish using low frequency based systems in shallow reverberant environments. This will be a subject for future work.

ACKNOWLEDGMENT

This work was supported in part by the Research Division Director's Fund.

Approved:

A.H. Chung
Manager
Operations Research Department

N. Anyas-Weiss
Manager
Divisional Projects Department

for J.E. Pagel
Manager
Environmental Sciences Department

Submitted:

N. Ramani
Engineer
Advanced Computational Methods Unit

W.G. Hanson
Research Program Scientist
Divisional Projects Department

P.H. Patrick
Associate Research Scientist
Biological Research Section

NR/WGH/PHP:cmh

REFERENCES

Mallat, S.G., (1989), "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 674-693.

Patrick, P.H., Ramani, N., Hanson, W.G., and Anderson, H.L., (1991), "The Potential of a Neural Network Based Sonar System in Classifying Fish", *Proceedings of the IEEE Conference on Neural Networks for Ocean Engineering*", Washington, August 15-17, 1991.

Ramani, N., Hanson, W.G., Patrick, P.H., and Anderson, H.L., (1991a), "Fish Identification from Sonar Echoes - Preprocessing and Parallel Networks", *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, Seattle, July 23-26, 1991.

Ramani, N., Hanson, W.G., and Patrick, P.H., (1991b), "Use of Neural Networks for Fish Identification from Sonar Echoes - Preliminary Results", Ontario Hydro Research Division Report No. 91-226-K.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., (1986), "Learning Internal Representations by Error Propagation", in D.E. Rumelhart and J.L. McClelland (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge: MIT Press, pp. 318-362.

FIGURE 1

Sample Echoes - 4 Different Species

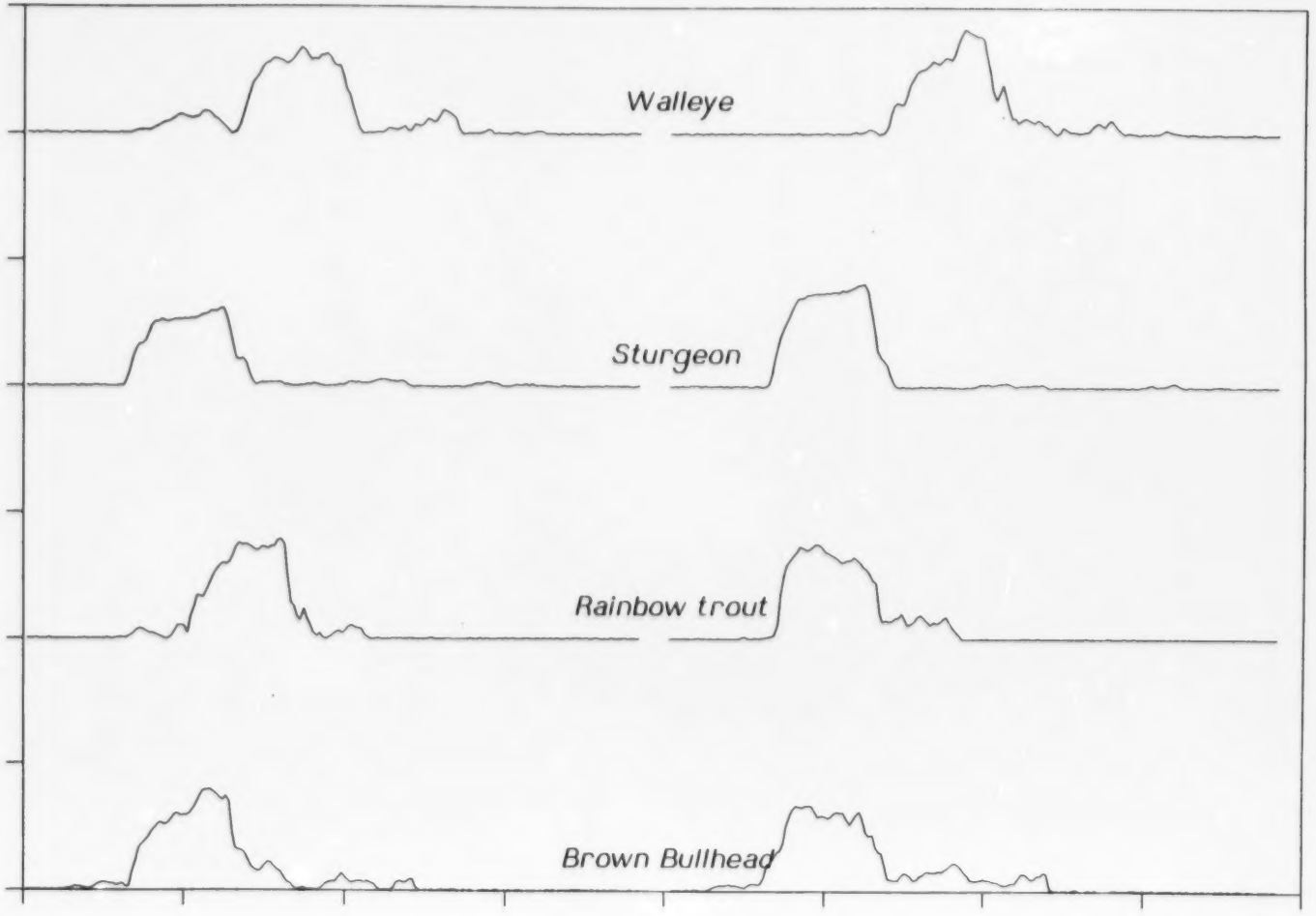

FIGURE 2

Sample Echoes - Walleye

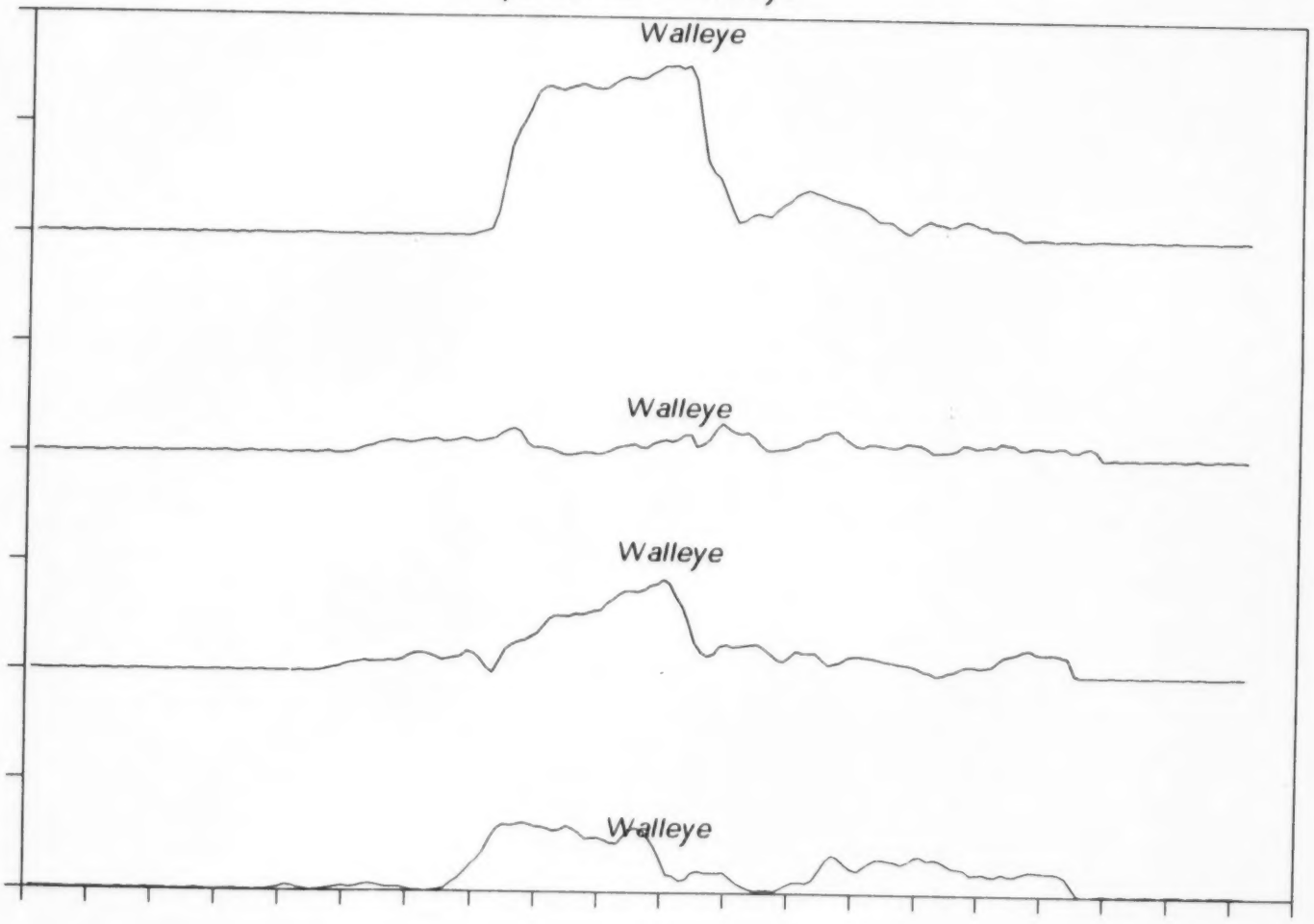

TABLE 1

ORIGINAL DATA - ONE SUBSET USED FOR TRAINING

Input Layer Size: 192, Output Layer Size: 9

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set[1]	Middle Layer Size					
	5	10	15	20	25	30
A:B	60	81	80	81	81	79
A:C	52	78	77	75	72	74
B:C	73	85	83	81	82	83
B:A	74	82	79	80	83	83
C:A	65	75	72	72	69	70
C:B	76	85	79	79	79	79

[1] For example, the networks in the first row are trained on subset A and tested on subset B.

TABLE 2

ORIGINAL DATA - TWO SUBSETS USED FOR TRAINING

Input Layer Size: 192, Output Layer Size: 9

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set[2]	Middle Layer Size					
	5	10	15	20	25	30
(A+B) : C	70	81	80	81	85	83
(B+C) : A	74	83	80	82	81	77
(C+A) : B	77	88	83	87	85	87

[2] For example, the networks in the first row are trained on subsets A and B, and tested on subset C.

TABLE 3

ORIGINAL DATA - FOUR LAYERED NETS

Input Layer Size: 192, Output Layer Size: 9

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set	Middle Layer Sizes					
	5,5	10,5	10,10	15,5	15,10	15,15
(A+B) : C	81	74	76	76	76	78
(B+C) : A	76	76	79	82	78	84
(C+A) : B	81	82	85	83	84	82

TABLE 4

WAVELET-BASED APPROXIMATION

Input Layer Size: 48, Output Layer Size: 9

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set	Middle Layer Size					
	5	10	15	20	25	30
(A+B) : C	81	85	87	83	83	86
(B+C) : A	83	80	83	80	81	79
(C+A) : B	87	85	83	85	85	86

TABLE 5

PEAK/TROUGH DETECTOR

Input Layer Size: 80, Output Layer Size: 9

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set	Middle Layer Size					
	5	10	15	20	25	30
(A+B) : C	79	84	81	84	84	82
(B+C) : A	88	90	92	91	92	90
(C+A) : B	87	88	88	88	86	87

TABLE 6

NETWORKS IN PARALLEL[3]

SUCCESS RATES (Percent correctly identified on test set)

Training set: Test set	Middle Layer Size					
	5	10	15	20	25	30
(A+B) : C	90	93	95	94	93	93
(B+C) : A	96	93	96	94	95	95
(C+A) : B	97	95	96	94	94	94

[3]Corresponding networks in Tables 4 and 5 in parallel.

APPENDIX A

THE NEURAL NETWORK APPROACH

The neural network approach is a biologically motivated statistical method for mapping between vector spaces. It has been proven that any arbitrary continuous mapping between two vector spaces can be realized using a neural network. In recent years, the approach has been used successfully in many pattern recognition applications, where a neural network maps an input pattern vector into an output decision vector. A distinctive feature of a neural network is that there is no algorithmic analysis of the problem, but the network is 'trained' by showing it samples of the desired mapping. Neural networks are particularly useful in cases where the data is noisy, as long as a statistically valid set of example data is available for training.

A neural network consists of an array of interconnected elementary processors (or neurons). The most popular model is called a multi-layer perceptron, where the network is arranged in layers, consisting of an input layer, one or more hidden layers and an output layer (see Figure A.1). Each neuron in a layer is connected to every neuron in the neighboring layer with distinct interconnection weights. A network performs an input-output mapping as follows. A signal representing the input vector is presented to the input layer. The activation level of each neuron in the other layers is a simple function of the weighted sum of the activation levels of the neurons in the prior layer. That is, the activation level o_j of neuron j is given by

$$o_j = f(net_j) \qquad \text{(A.1)}$$

where net_j is the net input to neuron j and is given by

$$net_j = \sum_i w_{ji} o_i \qquad \text{(A.2)}$$

where o_i is the output of neuron i in the prior layer, w_{ji} is the weight of the interconnection from neuron i to neuron j, and the summation is over all neurons in the prior layer. Note that, by definition, if neuron j is in the input layer, o_j is simply the input assigned to it. The activation levels of the neurons in the output layer represent the output vector. The hidden layer represents the key to the proper mapping by capturing functional representations that are not obvious in the input vector. These networks are also often called feedforward networks, because information propagates forward (layer by layer) through the network without any feedback.

The network starts with random weights and the weights are adapted, ie, the network is 'trained' to realize the appropriate mapping by showing it examples of the proper mapping. The commonly used adaptation algorithm is called the back-propagation algorithm (Rumelhart, Hinton and Williams, 1986) and is outlined below. It uses the gradient descent method to minimize the sum of the squared error at the output layer between the actual output and the expected (or target) output. That is, the cost function is given by

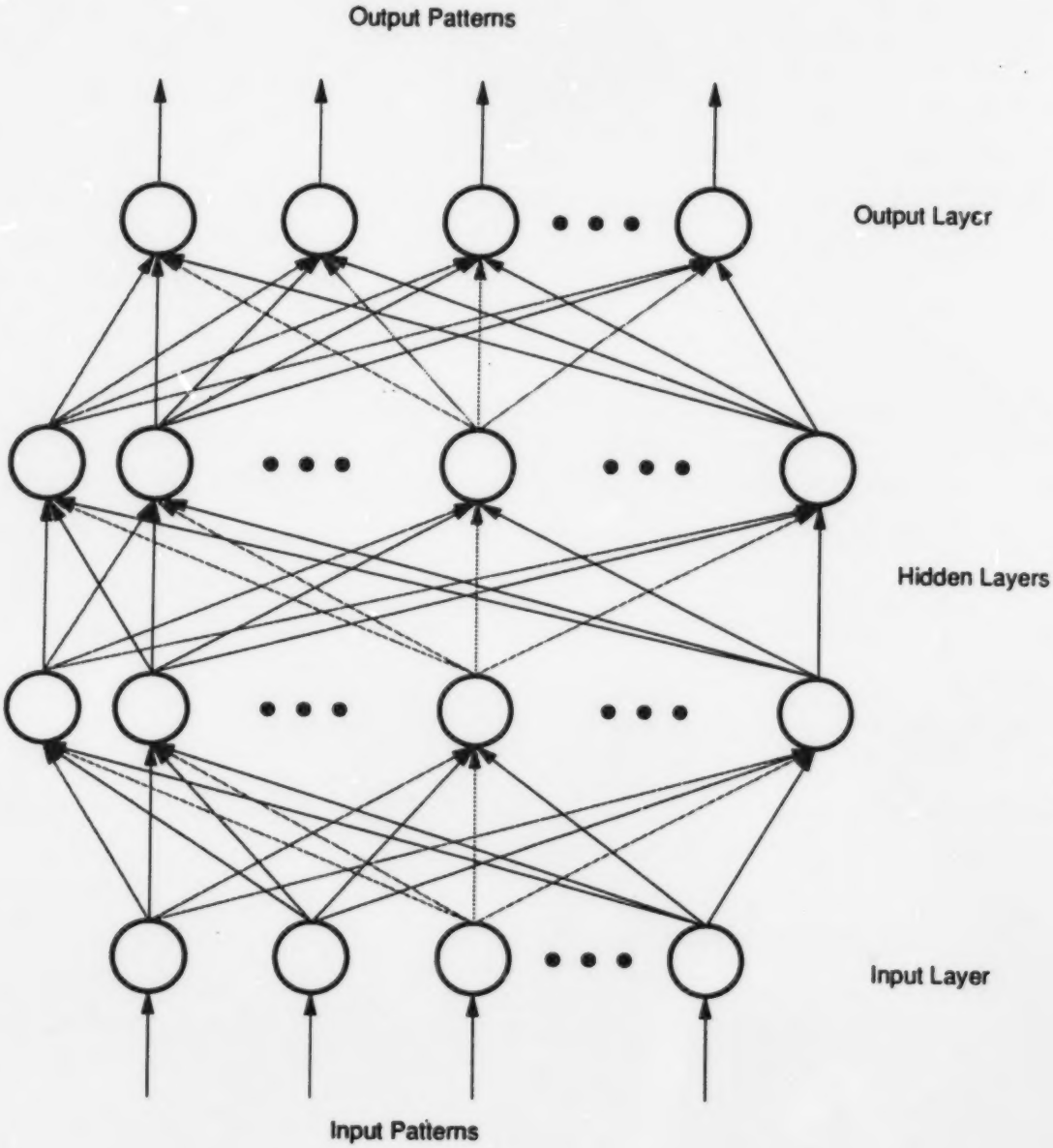

FIGURE A.1

A MULTILAYER FEEDFORWARD NETWORK

$$E = \frac{1}{2} \sum_j (t_j - o_j)^2 \qquad \text{(A.3)}$$

where the summation is over all the neurons in the output layer. According to the gradient descent rule for minimizing the cost function, the required weight change is given by (using a learning rate η)

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \qquad \text{(A.4)}$$

$$= -\eta \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \qquad \text{(A.5)}$$

$$= \eta \delta_j \frac{\partial net_j}{\partial w_{ji}} \qquad \text{(A.6)}$$

where,

$$\delta_j = -\frac{\partial E}{\partial net_j} \qquad \text{(A.7)}$$

$$= -\frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \qquad \text{(A.8)}$$

The first factor on the right hand side in Equation (A.8) depends upon the location of the neuron j. If the neuron is in the output layer, it is easy to see from Equation (A.3) that

$$\frac{\partial E}{\partial o_j} = -(t_j - o_j) \qquad \text{(A.9)}$$

If neuron j is not in the output layer, one can write

$$\frac{\partial E}{\partial o_j} = \sum_k \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial o_j} \qquad \text{(A.10)}$$

$$= -\sum_k \delta_k \frac{\partial net_k}{\partial o_j} \qquad \text{(A.11)}$$

where the summation is over all the neurons in the subsequent layer. One can see that the gradient of the error propagates backwards from the output layer giving rise tc the algorithm's name - 'back-propagation'.

The second factor on the right hand side of Equation (A.8) depends on the activation function f(x), which is usually chosen to be the sigmoidal function

$$f(x) = 1 / (1 + \exp(-x)) \qquad \text{(A.12)}$$

although the use of other functions have been reported in the literature. For the sigmoidal activation function, this term is given by

$$\frac{\partial o_j}{\partial net_j} = o_j (1 - o_j) \qquad \text{(A.13)}$$

From Equation (A.2), it can be seen that the last term on the right hand side in Equation (A.6) is given by

$$\frac{\partial net_j}{\partial w_{ji}} = o_i \qquad \text{(A.14)}$$

Combining Equations, (A.6, A.8, A.9, A.11, A.13, and A.14), the weight change algorithm can be summarized as follows:

$$\Delta w_{ji} = \eta \, \delta_j \, o_i \qquad \text{(A.15)}$$

$$\delta_j = (t_j - o_j) \, o_j (1 - o_j) \qquad \text{(A.16a)}$$

$$\delta_j = \left(\sum_k \delta_k w_{kj}\right) o_j (1 - o_j) \qquad \text{(A.16b)}$$

Equation (A.16a) is used if neuron j is in the output layer and Equation (A.16b) if it is not.

To improve convergence, Rumelhart, et al suggest the use of momentum term as well, so that the weight change at iteration n+1 is given by

$$\Delta w_{ji}(n+1) = \eta \, \delta_j \, o_i + \alpha \, \Delta w_{ji}(n) \qquad \text{(A.17)}$$

where α is the momentum factor.

In practice, weight changes are accumulated as each sample (input-output pair) is presented to the network, and changes are made only at the end of one cycle, ie after all the samples have been presented once. The above algorithm, with some minor modifications, is used in this study.